Complete genome sequence of methicillin-sensitive Staphylococcus aureus containing a heterogeneic staphylococcal cassette chromosome element

Citation for published version:
Li, D, Chu, Y, Ren, L, Li, X, Yuan, L, Kang, Y, Zhang, W, Yang, Y, Wang, X, Baillie, JK, Yu, J & Gao, Z 2013, 'Complete genome sequence of methicillin-sensitive Staphylococcus aureus containing a heterogeneic staphylococcal cassette chromosome element', *Science China Life Sciences*, vol. 56, no. 3, pp. 268-274. https://doi.org/10.1007/s11427-013-4453-9

Digital Object Identifier (DOI):
10.1007/s11427-013-4453-9

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Science China Life Sciences

SCIENCE CHINA
Life Sciences

• RESEARCH PAPER •

March 2013 Vol.56 No.3: 268–274
doi: 10.1007/s11427-013-4453-9

Complete genome sequence of methicillin-sensitive *Staphylococcus aureus* containing a heterogeneic staphylococcal cassette chromosome element

LI DeZhi[1,2†], CHU YaNan[3,4†], REN LuFeng[3,4], LI XinGang[3,4], YUAN Lina[3,4], KANG Yu[3], ZHANG Wei[5], YANG Yu[6], WANG XuMin[3], BAILLIE J. Kenneth[7,8], YU Jun[3,4*] & GAO ZhanCheng[1*]

[1]*Department of Respiratory and Critical Care Medicine, Peking University People's Hospital, Beijing 100044, China;*
[2]*Department of Respiratory Medicine, Shandong Provincial Hospital, Shandong University, Jinan 250021, China;*
[3]*CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100029, China;*
[4]*The DNA Sequencing Technologies R&D Center, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100029, China;*
[5]*Department of Respiratory Medicine, First Affiliated Hospital of Nanchang University, Nanchang 330006, China;*
[6]*Chinese Academy of Inspection and Quarantine, Beijing 100025, China;*
[7]*Division of Genetics and Genomics, The Roslin Institute, University of Edinburgh, Roslin, EH25 9RG, UK;*
[8]*Department of Critical Care Medicine, University of Edinburgh, Edinburgh, EH16 4TJ, UK*

Received October 30, 2012; accepted January 23, 2013

Staphylococcus aureus is a common human bacterium that sometimes becomes pathogenic, causing serious infections. A key feature of *S. aureus* is its ability to acquire resistance to antibiotics. The presence of the staphylococcal cassette chromosome (SCC) element in serotypes of *S. aureus* has been confirmed using multiplex PCR assays. The SCC element is the only vector known to carry the *mecA* gene, which encodes methicillin resistance in *S. aureus* infections. Here, we report the genome sequence of a novel methicillin-sensitive *S. aureus* (MSSA) strain: SCC-like MSSA463. This strain was originally erroneously serotyped as methicillin-resistant *S. aureus* in a clinical laboratory using multiplex PCR methods. We sequenced the genome of SCC-like MSSA463 using pyrosequencing techniques and compared it with known genome sequences of other *S. aureus* isolates. An open reading frame (CZ049; AB037671) was identified downstream of *attL* and *attR* inverted repeat sequences. Our results suggest that a lateral gene transfer occurred between *S. aureus* and other organisms, partially changing *S. aureus* infectivity. We propose that *attL* and *attR* inverted repeats in *S. aureus* serve as frequent insertion sites for exogenous genes.

***Staphylococcus aureus*, staphylococcal cassette chromosome mec (SCC-mec), genomics, pyrosequencing**

Citation: Li D Z, Chu Y N, Ren L F, et al. Complete genome sequence of methicillin-sensitive *Staphylococcus aureus* containing a heterogeneic staphylococcal cassette chromosome element. Sci China Life Sci, 2013, 56: 268–274, doi: 10.1007/s11427-013-4453-9

Staphylococcus aureus is a common human pathogen that can cause serious infections such as necrotizing pneumonia and sepsis. With the emergence and spread of methicillin-resistant *S. aureus* (MRSA) and vancomycin- and linezolid-resistant strains [1–3], a decreasing number of antibiotics is effective against this pathogen. It is therefore imperative to investigate the molecular basis and mechanisms of virulence and drug resistance in *S. aureus*.

One distinct feature of *S. aureus* is its ability to acquire

†Contributed equally to this work
*Corresponding author (email: zcgao@bjmu.edu.cn; junyu@big.ac.cn)

antibiotics resistance. Staphylococcal cassette chromosome (SCC) elements are currently the only vectors described for *mecA*, a gene that encodes methicillin resistance in *Staphylococcus* species. A lateral transfer event involving SCCmec was reportedly responsible for the acquired resistance [4]. The SCCmec element was originally thought to exist only as a complete sequence, but SCC elements have recently been detected in methicillin-sensitive *S. aureus* (MSSA) strains, suggesting loss of the portion of the SCC element that confers methicillin resistance [5]. High-throughput genome sequencing methods, especially those using next-generation platforms, afford the unprecedented ability to decipher sequence variations associated with the mechanism of resistance. In a previous study, the methicillin-susceptible strain MSSA476 was found to contain a novel SCCmec-like element (designated as SCC476). The SCCmec-like element is integrated into the same chromosome site as SCCmec elements in MRSA strains, but encodes a putative fusidic acid resistance protein [6]. These findings provide useful information regarding the molecular basis of SCCmec-like elements in MSSA.

In an earlier study [7], we discovered that the SCCmec III type-specific fragment open reading frame (ORF) CZ049 (AB037671; MRSA85/2082), previously considered unique and specific to SCCmec III in MRSA strains, is also present in the MSSA463 strain. This suggested that either fragment transfer or another structural change occurred during evolution of the strain [7]. In the study reported here, we used genomic techniques to determine key features of the SCC-like MSSA463 strain to gain new insights into virulence mechanisms of *S. aureus*.

1 Materials and methods

1.1 Ethics statement

The institutional review boards and ethics committees at First Affiliated Hospital of Nanchang University and Peking University People's Hospital approved the study protocol. Written informed consent was obtained from the patient from whom the bacterial strain was acquired prior to the beginning of this study. The study was conducted in accordance with the Declaration of Helsinki.

1.2 The bacterial strain

The MSSA463 strain with the SCCmec III type-specific fragment ORF CZ049 (AB037671) (MRSA85/2082) was originally designated as SCC-like MSSA. It was isolated from a 74-year-old patient in Nanchang, Jiangxi Province, China, who was diagnosed with acute exacerbation of chronic obstructive pulmonary disease, bronchiectasis, and community-acquired pneumonia. The strain was susceptible to most commonly-used antibiotics, including amoxicillin/CA, ampicillin/sulbactam, cefotaxime, ceftriaxone, ceftriaxone-sodium, ciprofloxacin, clindamycin, erythromycin, fusidic acid, gentamicin, imipenem, levofloxacin, linezolid, moxifloxacin, oxacillin, teicoplanin, tetracycline, trimethoprim/sulfa, and vancomycin. It was, however, resistant to ampicillin and penicillin G (Table S1).

1.3 Genome sequencing and assembly

We used a shotgun approach in conjunction with pyrosequencing [8] on a Roche/454 GS FLX platform (Roche Applied Science, Branford, US) to obtain 70-fold genomic coverage of MSSA463. The resulting sequence data was assembled using Newbler, an assembly program provided by the manufacturer. After refining the assembly to eliminate ambiguity, sequence gaps were closed by sequencing a series of PCR products. We used Glimmer 3.02 to predict open reading frames (ORFs) [9], which were further annotated based on the non-redundant (nr) database using the Basic Local Alignment Search Tool (BLAST) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) with an *E*-value cut-off of 10^{-5}. Proteins without homologs in the nr database were termed hypothetical proteins; those with homologs of unknown function were designated as conserved hypothetical proteins. Further analyses of the genome sequence were carried out using other databases, including UniProtKB/Swiss-Prot (http://www.ebi.ac.uk/uniprot/, *E*-value cut-off of 10^{-10}) [10], Kyoto Encyclopedia of Genes and Genomes (KEGG) (Release50; *E*-value cut-off of 10^{-10}) [11], Clusters of Orthologous Groups (COG) (*E*-value cut-off of 10^{-10}), Interpro (Interproscan 4.3, Release16.0) [12], and Gene Ontology (GO) [13] databases (Table S2).

1.4 Comparative genomics

We compared the MSSA463 genome sequence with related genome sequences using the Artemis Comparison Tool [14]. Orthologous proteins were identified as "reciprocal best matches" using FASTA followed by manual curation. Genome sequences used for comparison were MSSA476, MRSA252 (GenBank BX571856.1), Mu50 (GenBank BA000017.4), N315 (GenBank BA000018.3), and USA300_FPR3757 (GenBank CP000255.1).

2 Results

2.1 Genome assembly and annotation

Using the GS *de novo* assembler Newbler [8], we performed a *de novo* assembly of the initial reads (average read length of 352 bp and 70-fold coverage; Table S3), which yielded 33 contigs, 22 of which were longer than 500 kb. The contigs covered 98% of the genome (Table S3).

After gap closure, the final MSSA463 genome sequence was 2771498 bp long and contained 2563 predicted ORFs, with a gene density of 0.92 per kilobase and a simple repeat representation of 0.93% over the entire length. The low G+C content observed 32.45% is typical for Firmicutes. The longest ORF in the genome was 28236 bp long; the average gene length was 903 bp. The genome contained 5 rRNA operons, 59 tRNA ORFs, and 29 small RNA coding sequences (Table S4, Figure 1).

We used the COG database for gene annotation. Each COG cluster includes proteins that are inferred to be orthologs, i.e., direct evolutionary counterparts (Figure S1). There were 208 orthologs classified into the unknown functions category and 232 orthologs for which only general functions were predicted (Figure S1). Gene Ontology categories for MSSA463 ORFs are shown in Figure S2.

2.2 Multiple alignments of *S. aureus* genomes

Using the Artemis Comparison Tool [14], we aligned the MSSA463 genome sequence to four other *S. aureus* genome assemblies: MSSA476, a community-acquired MSSA; USA300, a community-acquired MRSA; N315, a hospital-acquired MRSA; and Mu50, a vancomycin intermediate sensitive strain or VISA. All of these genomes were ~2.8 Mb in size and had a G+C content of 32%. Overall, 2235 genes (87.2%) in the MSSA463 genome were also present in all four of the other strains. Eighty-seven genes (3.4%) were found in only three of the four reference strains, an additional 40 genes (1.6%) were present in only two of the reference strains. Fifty-six genes (2.2%) were shared with only one reference genome. In total, we identified 145 genes that were unique to MSSA463 (Figure S3).

2.3 Analysis of mobile elements

Using reference strains MSSA476 and MRSA252, we examined the MSSA463 genome for variations in genomic islands and mobile regions, the locations of genes involved in virulence or drug resistance [15]. Although we did not find the typical SCC structure in MSSA463 (Table 1), temperate bacteriophages were present; this was expected, as most known *S. aureus* strains harbor more than one bacteriophage. MSSA463 was discovered to possess the same Saα-vSaβ island including four staphylococcal exotoxin genes (*spl*, *lukDE*, *hysA*, and *bsa*) present in six other *S. aureus* strains (MSSA476, MRSA252, N315, Mu50, MW2, and NCTC8325). When we analyzed pathogenicity islands, we found the pathogenicity island SaPI6 in MSSA463 and MSSA476, but not in 8325, COL, USA300, or Mw2. Similarly, two pathogenicity islands, SaPI4 and SaPI1028, were only detected in MRSA252. Insertion sequences were also found in the *S. aureus* strains. For example, IS1272 and IS431 were present in MRSA252 and IS431 was present in MSSA476, whereas, no transposons were detected in

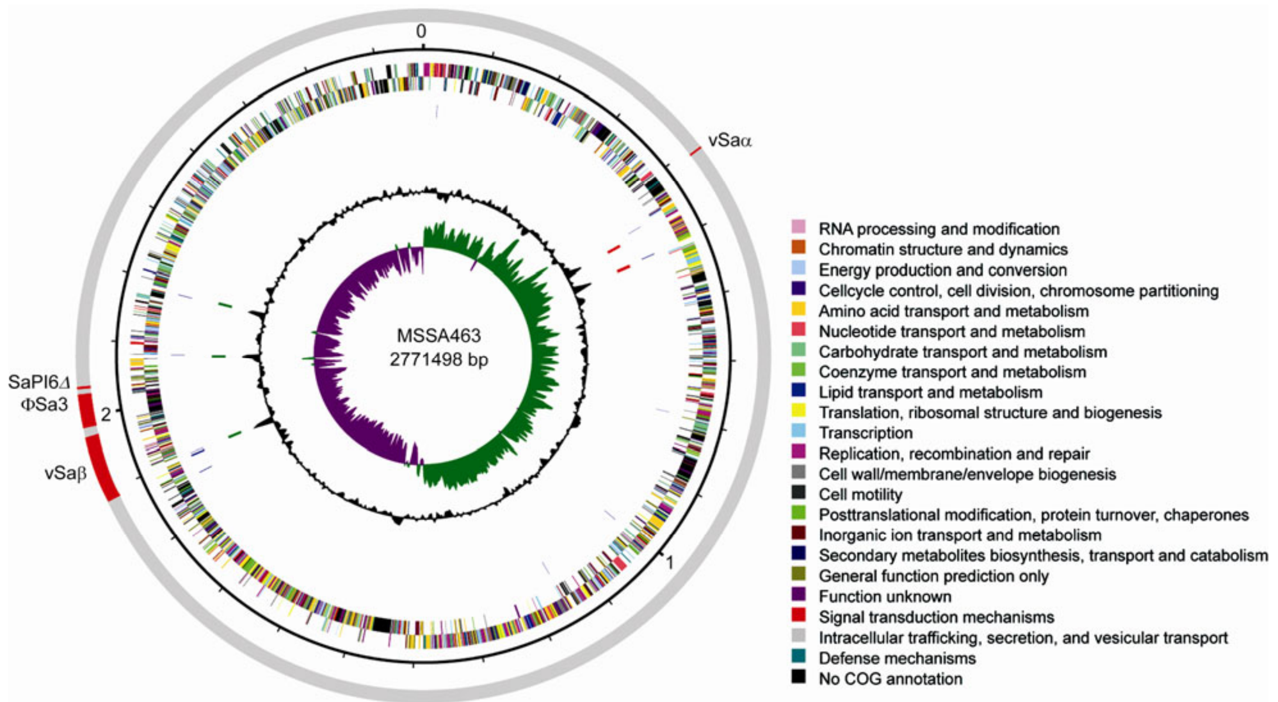

Figure 1 Schematic circular diagram of the SCC-like MSSA463 chromosome. Where appropriate, categories are shown as concentric circles representing coding strands. From the outside to inside: the colored segments on the gray outer ring represent genomic islands and horizontally acquired DNA. Inside the gray outer circle, the second circle represents the nucleotide position in Mb. The third and fourth circles show open reading frames (ORFs) on the plus and minus strands, respectively. Different colors represent different Clusters of Orthologous Groups (COG) functions. The color coding for coding sequences (CDS) is listed to the right of the circular diagram. The fifth circle indicates the location of predicted tRNAs. The sixth and seventh circles represent rRNA predicted on the plus and minus strands, respectively. The eighth and ninth circles show G+C content and G+C skew, respectively (>0%, green; <0%, purple).

Table 1 Mobile elements of Staphylococcus in MSSA463, MSSA476 and MRSA252

Mobile element	Length (kbs)	Gene of virulence or drug resistance	MSSA463	MSSA476	MRSA252
SCC					
Type II SCCmec	~53.0	*mecA*, *mecR1*, *mecI*	−	−	+
SCC476	~23.0	*Far*	−	+	−
Bacteriophage					
ΦSa2	~45.8	*lukSF-PV*, *Φ12*	−	−	+
ΦSa3	~45.0	*sak*, *sep*, *sea*, *seg2*, *sek2*	+	+	+
ΦSa4	~44.7	none	−	+	−
Genomic islands					
νSaα	~3.8	*Set*	+	+	+
νSaβ	~155.8	*spl*, *lukDE*, *hysA*, *bsa*, *extoxin*	+	+	+
Pathogenicity islands					
SaPI4	~15.1	none	−	−	+
SaPI1028	~15.6	none	−	−	+
SaPI6⊿	~3.14	none	+	+	−
SaPI15305	~17.0	*aad*, *fosB*	−	−	−
Insertion sequences					
IS1272	~2.0	none	−	−	+
IS431	~0.8	none	−	+	+
ISX	~1.1	none	−	−	+
Transposons					
Tn554	~6.7	*ermA*, *spc*	−	−	+
Tn552	~6.5	*blaZ*	−	−	+
Tn916-like	~7.6	none	−	−	+

MSSA463 and MSSA476. In general, a large number of mobile elements were present in MRSA than in MSSA strains (e.g., neither IS1272 nor IS431 was detected in MSSA strains).

2.4 Pathogenicity-associated coding sequences (CDS) and cell wall/membrane-associated resistance genes

We used the KEGG database for further functional gene prediction in MSSA463, with a focus on disease-related CDS. Multiple genes associated with infectious disease were identified in the MSSA463 genome. These included *fnbB* and *fnbA*, which are involved in bacterial invasion of epithelial cells, K08303 (a putative protease), and *ureA*, *ureB*, and *ureC*, which are associated with *Helicobacter pylori* infections of epithelial cells. Other identified genes included *luxS*, associated with the *Vibrio cholerae* pathogenic cycle, and arginases such as *rocF* (K0416; E3.5.3.1) and *arg*, implicated in eukaryotic amebiasis (Table S5).

Virulence and drug-resistance proteins are generally thought to be located on the cell wall or membrane [16]. Some exogenous pathogenic factors, such as adhesion-related *SdrD*, beta-channel forming cytolysin, and clumping factor B, were located in ORFs of MSSA463. Although MSSA463 is sensitive to methicillin, we found some resistance protein-coding genes, including bicyclomycin resistance protein TcaB and chloramphenicol resistance protein. These results are consistent with the phenotypic characteristics of the strain.

2.5 Phylogenetic relationship of MSSA463 and its SCC-like element to other *S. aureus* strains

We used a web-based tool, Interactive Tree Of Life or iTOL, to generate and display phylogenetic trees [17] (Figure S3). In this tree, MSSA463 is phylogenetically closest to MRSA252 (gi 49240382; Figure 2).

Based on the generated phylogeny, we selected several strains, including MSSA463, MRSA85/2082, MSSA476, and MRSA252 [6] for a more detailed comparative analysis of the SCC-like element (Figure S4). All SCCmec elements were typical, integrating at exactly the same site across all of strains. The inverted repeat sequences found in MRSA strains were also present in MSSA463; in contrast, SCCmec components (including the conserved site-specific recombination enzymes *ccrA* and *ccrB*) were completely missing from MSSA463. The specific ORF CZ049 sequence (for the basis of diagnostic detection using PCR assays) was present in MSSA463, but was located 1686 bp downstream of *attL* and *attR* inverted repeat sequences , rather than between them as in other strains.

3 Discussion

Following the sequencing of the first *S. aureus* N315 strain, 22 additional *S. aureus* genomes have been deposited in the NCBI genome database (http://www.ncbi.nlm. nih.gov/ genomes/lproks.cgi). Because the strains investigated in our study, corresponding to the 22nd genome sequence added to

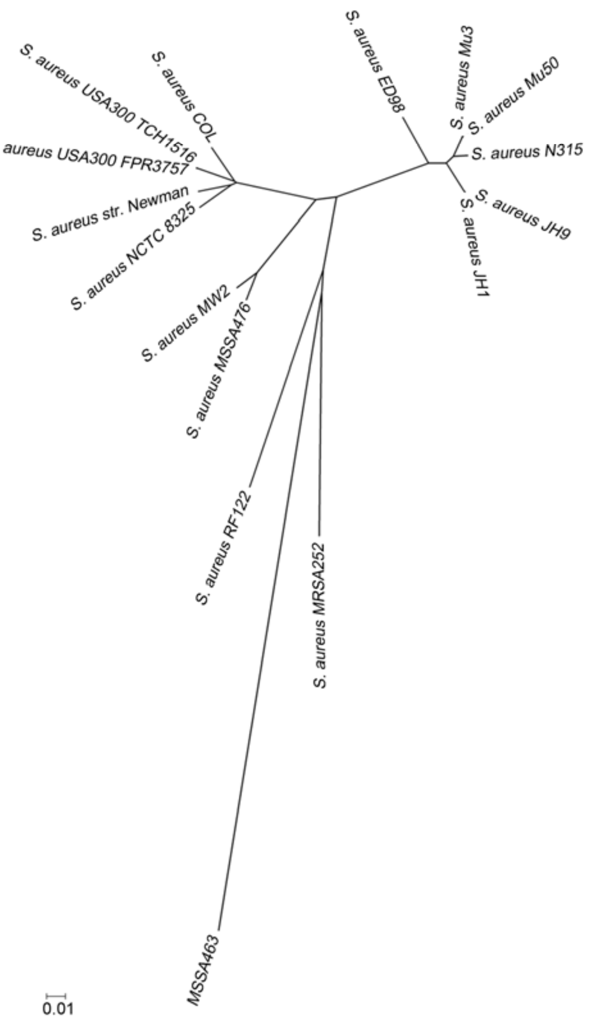

Figure 2 A phylogeny based on all sequenced *Staphylococcus aureus* strains. The split decomposition tree was constructed from *S. aureus* total genome sequences using SPLITSTREE Version 3.1.

the database, is not a typical SCC *S. aureus*, it has been designated as SCC-like MSSA463.

MSSA463 is pathogenic. Many pathogenic bacteria can invade both phagocytic and non-phagocytic cells and colonize them intracellularly, with subsequent dissemination to other cell types [18–21]. Introduction into non-phagocytic host cells, such as epithelial cells, occurs via two different mechanisms, i.e., the "zipper" and "trigger" models. Pathogenic bacteria that enter their hosts using the zipper model include *Listeria*, *Streptococcus*, and *Yersinia*. The presence of *fnbB* and *fnbA* in MSSA463 suggests that *S. aureus* uses these zipper model genes to initiates the infection cascade. The abundant presence of other pathogenic genes in MSSA463, including *ureABC*, *luxS*, *rocF*, and *arg*, is evidence for similar pathways shared in common with *H. pylori*, *V. cholerae*, and infectious amoeba [16,22–24]. These shared pathway raise the possibility of frequent horizontal gene transfer between these pathogens as well as between prokaryotic and eukaryotic pathogens [25,26].

In addition to its pathogenicity, MSSA463 is drug- and antibiotic-resistant. With the development of genome sequencing technology and the consequent accumulation of genome-scale information, our understanding of drug resistance has significantly improved. Katayama et al. [27] were the first to demonstrate that the MRSA methicillin-resistance gene is carried by a novel genetic element, SCCmec, which is integrated into and excised from the *S. aureus* chromosome through the mediation of a unique set of recombinase genes: *ccrA* and *ccrB*. They were also the first to define the structure of SCCmec elements. SCCmec elements are currently classified into different groups based on the nature of *mec* and *ccr* gene complexes, and are further classified into different subtypes according to characteristics of their "junkyard" regions [28,29]. In MRSA strains, two SCCmec elements are located within the left and right boundaries of *attL* and *attR* inverted repeat sequences, respectively. It is typically believed that these two SCCmec elements co-exist, and that the lateral transfer of this element as a whole from other bacteria gave rise to methicillin resistance in *S. aureus*. The presence of SCC476 elements in MSSA476, however, suggests that other forms of the SCCmec element may also exist in MSSA strains. The existence of these elements may be a common phenomenon in *S. aureus*, or may simply be the result of the loss of other elements from MRSA strains. While this theory needs to be further explored, these findings increase our understanding of the molecular details of bacterial resistance [7].

Our discovery of an SCC-like sequence in MSSA476 has important implications for drug resistance diagnosis. Previous preliminary molecular epidemiology studies of *S. aureus* in lower respiratory tract infections [7] used multiplex PCR assay methods to characterize and subtype SCCmec elements. The presence of ORF CZ049 was considered to be specific to SCCmec III strains such as MRSA85/2082 [28]. In MRSA85/2082 strain, ORF CZ049 is located between *attL* and *attR* inverted repeat sequences, as defined by the SCCmec structure. Although this SCCmec III type-specific fragment is also present in other SCCmec type strains, including MRSA strain JCSC3624(WIS) [30] and MSSA strain MSSA463, considered to be a unique marker for SCCmec III in MRSA strains [28]. Interestingly, ORF CZ049 is also found in MSSA463.

One possible explanation for the unusual location of CZ049 in MSSA463 is that MSSA463 is derived from an MRSA strain in which ORF CZ049 and the SCC element were in close proximity in the chromosome. Alternatively, the SCCmec element may have been duplicated upstream of ORF CZ049, with subsequent loss of the SCCmec element at the original location. A third possibility is that the abnormal location of CZ049 arose as a consequence of chromosome recombination events [31,32]. Whether the downstream location of ORF CZ049 is the result of a short element insertion or is instead due to secondary changes associated with loss of other SCCmec components, remains to be determined.

In summary, we used genome sequencing and comparative analysis to acquire novel insights into *S. aureus* pathology and resistance mechanisms and the infection cascades involved in its pathogenicity. The role of inverted repeat sequences in antibiotic resistance requires further study. As more non-classical SCC-like genome sequences are completed, we expect that a more complete understanding of the mechanisms underlying generation and function of antibiotic resistance genes in *S. aureus* will be developed.

This work was supported by the National High Technology Research and Development Program (2006AA02Z4A9), the National Science and Technology Major Project of Ministry of Science and Technology of China (2009ZX10004, 2012ZX10004206), and the National Natural Science Foundation of China (30971610, 30900053). Funding awarded to Ren LuFeng is from the Young Scientist Innovation Promotion Association of the Chinese Academy of Sciences.

1 Hiramatsu K, Aritaka N, Hanaki H, et al. Dissemination in Japanese hospitals of strains of *Staphylococcus aureus* heterogeneously resistant to vancomycin. Lancet, 1997, 350: 1670–1673
2 Vancomycin-resistant *Staphylococcus aureus*—New York, 2004. MMWR Morb Mortal Wkly Rep, 2004, 53: 322–323
3 Jones R N, Kohno S, Ono Y, et al. ZAAPS International Surveillance Program (2007) for linezolid resistance: results from 5591 Gram-positive clinical isolates in 23 countries. Diagn Micr Infec Dis, 2009, 64: 191–201
4 Pantosti A, Sanchini A, Monaco M. Mechanisms of antibiotic resistance in Staphylococcus aureus. Future Microbiol, 2007, 2: 323–334
5 Corkill J E, Anson J J, Griffiths P, et al. Detection of elements of the staphylococcal cassette chromosome (SCC) in a methicillin-susceptible (mecA gene negative) homologue of a fucidin-resistant MRSA. J Antimicrob Chemother, 2004, 54: 229–231
6 Holden M T, Feil E J, Lindsay J A, et al. Complete genomes of two clinical Staphylococcus aureus strains: evidence for the rapid evolution of virulence and drug resistance. Proc Natl Acad Sci USA, 2004, 101: 9786–9791
7 Li D Z, Chen Y S, Yang J P, et al. Preliminary molecular epidemiology of the *Staphylococcus aureus* in lower respiratory tract infections: a multicenter study in China. Chin Med J (Engl), 2011, 124: 687–692
8 Margulies M, Egholm M, Altman W E, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature, 2005, 437: 376–380
9 Delcher A L, Bratke K A, Powers E C, et al. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics, 2007, 23: 673–679
10 Bairoch A, Apweiler R. The SWISS-PROT protein sequence data bank and its new supplement TREMBL. Nucleic Acids Res, 1996, 24: 21–25
11 Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res, 2000, 28: 27–30
12 Mulder N J, Apweiler R, Attwood T K, et al. InterPro: an integrated documentation resource for protein families, domains and functional sites. Brief Bioinform, 2002, 3: 225–235
13 Harris M A, Clark J, Ireland A, et al. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res, 2004, 32: D258–261
14 Carver T J, Rutherford K M, Berriman M, et al. ACT: the Artemis Comparison Tool. Bioinformatics, 2005, 21: 3422–3423
15 Lindsay J A, Holden M T. *Staphylococcus aureus*: superbug, super genome? Trends Microbiol, 2004, 12: 378–385
16 Hsiao A, Liu Z, Joelsson A, et al. Vibrio cholerae virulence regulator-coordinated evasion of host immunity. Proc Natl Acad Sci USA, 2006, 103: 14542–14547
17 Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res, 2011, 39: W475–W478
18 Mostowy S, Cossart P. Cytoskeleton rearrangements during listeria infection: clathrin and septins as new players in the game. Cell Motil Cytoskel, 2009, 66: 816–823
19 Wang B N, Li S Y, Dedhar S, et al. Paxillin phosphorylation: bifurcation point downstream of integrin-linked kinase (ILK) in streptococcal invasion. Cell Microbiol, 2007, 9: 1519–1528
20 Matsumoto H, Young G M. Translocated effectors of Yersinia. Curr Opin Microbiol, 2009, 12: 94–100
21 Iwai H, Kim M, Yoshikawa Y, et al. A bacterial effector targets Mad2L2, an APC inhibitor, to modulate host cell cycling. Cell, 2007, 130: 611–623
22 Petri W A, Haque R, Mann B J. The bittersweet interface of parasite and host: Lectin-carbohydrate interactions during human invasion by the parasite *Entamoeba histolytica*. Annu Rev Microbiol, 2002, 56: 39–64
23 Gerhard M, Rad R, Prinz C, et al. Pathogenesis of *Helicobacter pylori* infection. Helicobacter, 2002, 7: 17–23
24 Montecucco C, Rappuoli R. Living dangerously: how Helicobacter pylori survives in the human stomach. Nat Rev Mol Cell Bio, 2001, 2: 457–466
25 Klotz M G, Loewen P C. The molecular evolution of catalatic hydroperoxidases: Evidence for multiple lateral transfer of genes between prokaryota and from bacteria into Eukaryota. Mol Biol Evol, 2003, 20: 1098–1112
26 Keeling P J, Palmer J D. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet, 2008, 9: 605–618
27 Katayama Y, Ito T, Hiramatsu K. A new class of genetic element, staphylococcus cassette chromosome mec, encodes methicillin resistance in *Staphylococcus aureus*. Antimicrob Agents Chemother, 2000, 44: 1549–1555
28 Zhang K Y, McClure J A, Elsayed S, et al. Novel multiplex PCR assay for characterization and concomitant subtyping of staphylococcal cassette chromosome mec types I to V in methicillin-resistant *Staphylococcus aureus*. J Clin Microbiol, 2005, 43: 5026–5033
29 Higuchi W, Takano T, Teng L J, et al. Structure and specific detection of staphylococcal cassette chromosome mec type VII. Biochem Bioph Res Co, 2008, 377: 752–756
30 Ito T, Ma X X, Takeuchi F, et al. Novel type v staphylococcal cassette chromosome mec driven by a novel cassette chromosome recombinase, ccrC. Antimicrob Agents Chemother, 2004, 48: 2637–2651
31 Robinson D A, Enright M C. Evolution of Staphylococcus aureus by large chromosomal replacements. J Bacteriol, 2004, 186: 1060–1064
32 Price L B, Stegger M, Hasman H, et al. *Staphylococcus aureus* CC398: host adaptation and emergence of methicillin resistance in livestock. Mbio, 2012, 3: e00305-00311

Supporting Information

Figure S1 Clusters of Orthologous Groups (COG) of protein categories of MSSA463. The numbers and the proportion of *Staphylococcus aureus* orthologs with various biological roles are shown. These genes are involved in cell cycle control; cell division; chromosome partitioning; cell motility; cell wall/membrane/envelope biogenesis; defense mechanisms; intracellular trafficking, secretion, and vesicular transport; post-translational modification; protein turnover; chaperones; signal transduction mechanisms; replication; recombination and repair; transcription; translation; ribosomal structure and biogenesis; amino acid transport and metabolism; carbohydrate transport and metabolism; coenzyme transport and metabolism; energy production and conversion; inorganic ion transport and metabolism; lipid transport and metabolism; nucleotide transport and metabolism; and secondary metabolite biosynthesis, transport, and catabolism. There are 208 orthologs annotated with unknown functions and 232 orthologs with only a predicted general function.

Figure S2 Gene Ontology categories for MSSA463 open reading frames. The bars show the numbers and the percentage of genes belonging to each functional category in Gene Ontology.

Figure S3 MUMmer-based homology alignment of whole genomes of MSSA463 with reference strains MSSA476 (A), USA300 (B), N315 (C), and Mu50 (D). The red lines indicate homologous genes at the same location. The blue lines indicate homologous genes at the opposite position.

Figure S4 A comparison of staphylococcal cassette chromosome (SCC) and SCC-like elements. Schematic diagrams of SCC-like MSSA463 (first schematic), the type III SCCmec of MRSA85/2082 (second schematic), non-mecSCC476 of MSSA476 (third schematic), and the type II SCCmec of MRSA252 (fourth schematic) are shown. Genes are marked with arrows in the direction of transcription. Genes with similarity in all elements are marked in the same color; genes with no similarity are white. Conservative regions in the SCC elements are joined by light blue shading. The dark blue boxes at the end of the SCC elements mark the left and right attachment sites (i.e., *attL* and *attR*), respectively. The complementary inverted repeats at the boundaries of the elements are marked with red arrows.

Table S1 The drug sensitivity of *Staphylococcus aureus* MSSA463

Table S2 Databases and their version used for the genome analysis

Table S3 Initial sequencing data

Table S4 General features of the *Staphylococcus aureus* MSSA463 genome

Table S5 Coding sequences (CDS) predicted to be involved in pathways associated with human diseases recorded in the Kyoto Encyclopedia of Genes and Genomes

The supporting information is available online at life.scichina.com and www.springerlink.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.